Dense RGB-D mapping of large scale environments for real-time localisation and autonomous navigation

Maxime Meilland, Patrick Rives, Andrew I. Comport

Dense RGB-D mapping of large scale environments for real-time localisation and autonomous navigation

Maxime Meilland, Patrick Rives and Andrew Ian Comport

***Abstract*— This paper presents a method and apparatus for building 3D dense visual maps of large scale environments for real-time localisation and autonomous navigation. We propose a spherical ego-centric representation of the environment which is able to reproduce photo-realistic omnidirectional views of captured environments. This representation is composed of a graph of locally accurate *augmented spherical panoramas* that allows to generate varying viewpoints through novel view synthesis. The spheres are related by a graph of 6 *d.o.f.* poses which are estimated through multi-view spherical registration. It is shown that this representation can be used to accurately localise a vehicle navigating within the spherical graph, using only a monocular camera for accurate localisation. To perform this task, an efficient direct image registration technique is employed. This approach directly exploits the advantages of the spherical representation by minimising a photometric error between a current image and a reference sphere. Autonomous navigation results are shown in challenging urban environments, containing pedestrians and other vehicles.**

I. INTRODUCTION

Acquiring 3D models of large scale environments is currently a key issue for a wide range of applications ranging from interactive personal guidance devices to autonomous navigation of mobile robots. In these applications it is important, not only for human operators but also for autonomous robots, to maintain a world map that holds a rich set of data including photometric, geometric and saliency information. It will be shown in this paper why it is advantageous to define an *ego-centric* representation of this information that allows fast model acquisition whilst maintaining optimal realism and accuracy.

An a-priori 3D model simplifies the localisation and navigation task since it allows to decouple the structure and motion estimation problems. Current state of the art approaches mostly rely on global 3D CAD models [10] that are based on tools and representations that have been developed mainly for texture mapped virtual reality environments. Unfortunately, these representations have difficulties in maintaining true photo-realism and therefore they introduce reconstruction errors and photometric inconsistencies. Furthermore, these models are complicated to acquire and often resort to heavy off-line modelling procedures. Whilst efforts are being made to use sensor acquisition systems that automatically acquire these classical virtual 3D models [7], it is suggested in this paper that they are not sufficient to precisely represent real-world data. Alternatively, it is proposed to use an ego-centric model [17] that represents, as close as possible, real sensor measurements.

A well known ego-centric representation model for camera sensors is the spherical panorama. Multiple cameras systems such as in [2] allow construction of high resolution spherical views via image stitching algorithms such as reviewed in [21]. However, contrary to virtual reality models, these tools have been developed mainly for qualitative photo-consistency but they rarely require 3D geometric consistency of the scene. This is mainly due to the fact that, in most cases, it is impossible to obtain 3D structure via triangulation of points when there is no or little baseline between images. Another approach is to use a central catadioptric omnidirectional camera [20] and warp the image plane onto a unit sphere using the model given in [8]. Unfortunately, that kind of sensor has a poor and varying spatial resolution and therefore is not well adapted to a visual memory of the environment. Furthermore, these approaches assume a unique center of projection, however, manufacturing such a system is still a challenging problem [14].

In order to take advantage of both 3D model based approaches and photometric panoramas it is possible to *augment* the spherical image with a depth image containing a range for each pixel. The multi-camera sensor proposed in [18] allows to acquire high resolution spherical images augmented with depth information. An augmented sphere then allows to perform novel view synthesis in a local domain in all directions [1][6][17].

These ego-centric models are, however, local and do not provide a global representation of the environment. This problem can be solved by considering multiple augmented spheres connected by a *graph* of poses that are positioned optimally in the environment. Simple spherical images positioned in the environment are already found in commercial applications such as Google Street View, and more recently in [13]. The easiest method for positioning spheres would be via a global positioning system (GPS). However, in urban environments this system fails easily due to satellite occlusion. Alternatively, the robot-centred representation introduced in [17][18] positions augmented views globally within a precise topological graph via accurate spherical visual odometry [6] and does not require any external sensor. The present paper extends this preliminary work, and demonstrates autonomous navigation results in challenging environments using only a monocular camera for localisation.

M. Meilland and P. Rives are with INRIA Sophia Antipolis Méditerranée, 2004 Route des Lucioles BP 93, Sophia Antipolis, France, {name.surname}@inria.fr

A.I. Comport is with CNRS, I3S Laboratory, Université Nice Sophia Antipolis, 2000 Route des Lucioles BP 121, Sophia Antipolis, France, comport@i3s.unice.fr

II. Real-time Ego-centric Tracking

The objective of this work is to perform real-time localisation and autonomous navigation using a known environment model (see Fig. 2). The essential part of this paper is therefore divided into two distinct but inter-related aspects:

- **Offline learning** - This phase consists in acquiring a 3D model of the environment and representing this information in an optimal manner for "on-line" localisation. It has been chosen to develop a learning approach that is also *efficient* so that, firstly, environments can be acquired rapidly and secondly, so that the approach may be used for online mapping in the near future. Essentially this involves filming, tracking and mapping the 3D environment (≈1Hz depending on the approach). The local ego-centric 3D model and the global graph learning are illustrated in Section III.
- **Online localisation and autonomous navigation** - The real-time phase involves estimating the 6 *d.o.f.* pose of a camera at frame-rate (here 45 Hz), onboard the vehicle. This phase must take into account efficient optimisation techniques that require a maximum amount of computation to be performed "off-line" during the learning phase. An accurate vehicle position should be provided to autonomously control the vehicle, as explained in Sections IV and V.

III. Spherical ego-centred model

An ego-centric 3D model of the environment is defined by a graph $\mathcal{G} = \{\boldsymbol{\mathcal{S}}_1, \ldots, \boldsymbol{\mathcal{S}}_n; \mathbf{x}_1, \ldots, \mathbf{x}_m\}$ where $\boldsymbol{\mathcal{S}}_i$ are *augmented spheres* that are connected by a minimal parametrisation $\mathbf{x}$ of each pose as:

$$\mathbf{T}(\mathbf{x}) = e^{[\mathbf{x}]_\wedge} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \in \mathbb{SE}(3), \tag{1}$$

where $\mathbf{x}^{ab} \in \mathbb{R}^6$ is the 6 *d.o.f.* twist between the sphere a and b (see Fig. 2) defined as:

$$\mathbf{x} = \int_0^1 (\boldsymbol{\omega}, \boldsymbol{\upsilon}) dt \in \mathfrak{se}(3), \tag{2}$$

which is the integral of a constant velocity twist which produces a pose $\mathbf{T}$. The operator $[.]_\wedge$ is defined as follows:

$$[\mathbf{x}]_\wedge = \begin{bmatrix} [\boldsymbol{\omega}]_\times & \boldsymbol{\upsilon} \\ \mathbf{0} & 0 \end{bmatrix}, \tag{3}$$

where $[.]_\times$ represents the skew symmetric matrix operator.

A. Augmented visual sphere

Each sphere is defined by the set

$$\boldsymbol{\mathcal{S}} = \{\boldsymbol{\mathcal{I}}_S, \boldsymbol{\mathcal{P}}_S, \boldsymbol{\mathcal{Z}}_S, \boldsymbol{\mathcal{W}}_S\}, \tag{4}$$

where:

- $\boldsymbol{\mathcal{I}}_S$ is the photometric spherical image. This image is obtained from the custom camera system presented in Section III-B by warping multiple images onto the sphere.
- $\boldsymbol{\mathcal{P}}_S = \{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$ is a set of evenly spaced points on the unit sphere where $\mathbf{q} \in S^2$. These points have been sampled uniformly on the sphere as in [17].

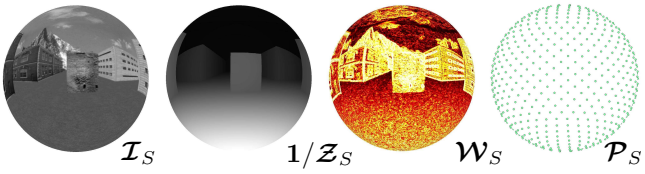


Fig. 1. Local representation: augmented sphere $\boldsymbol{\mathcal{S}}$ containing intensities $\boldsymbol{\mathcal{I}}_S$, depthmap $\boldsymbol{\mathcal{Z}}_S$, salliancy $\boldsymbol{\mathcal{W}}_S$ and a sampling $\boldsymbol{\mathcal{P}}_S$.

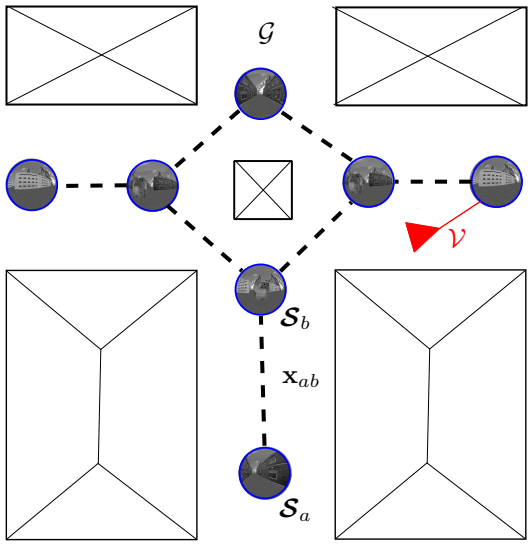


Fig. 2. Ego-centric representation: graph of spheres $\mathcal{G}$ allowing the localisation of a vehicle $\mathcal{V}$ navigating locally within the graph.

- $\boldsymbol{\mathcal{Z}}_S$ are the depths associated with each pixel which have been obtained from dense stereo matching. The 3D point is subsequently defined in the sphere as $\mathbf{P} = (\mathbf{q}, Z)$.
- $\boldsymbol{\mathcal{W}}_S$ is a saliency image which contains knowledge of good pixels to use for tracking applications. It is obtained by analysing the Jacobian of the warping function so that the pixels are ordered from best to worst in terms of how they condition the pose estimation problem (the interested reader can see [17] for more details).

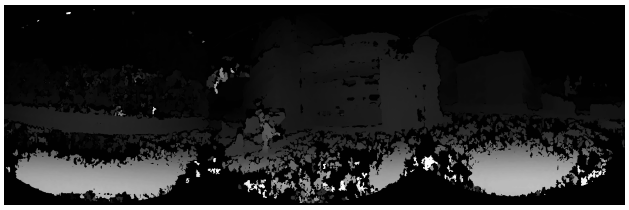
B. Spherical acquisition system

To acquire augmented visual spheres, a custom made multi-camera sensor is employed. This sensor, proposed in [18] is composed of six wide angle stereo cameras, placed in a ring configuration so that dense stereo matching [11] can be performed between each camera pair. Such a system allows to acquire high resolution visual spheres, with dense depth information for almost each pixels of the spherical image, as it can be seen on Figure 3.

C. Global sphere positioning

To accurately recover the position of the spheres of the graph with respect to one-another, a 6 *d.o.f.* multi-camera localisation model is used based on accurate dense localisation [6][17]. Considering $\boldsymbol{\mathcal{S}}^*$, an augmented sphere defined in Section III-B, the objective is to compute the pose between a reference sphere $\boldsymbol{\mathcal{S}}^*$ and the next one $\boldsymbol{\mathcal{S}}$. The localisation

(a) Spherical image.



(b) Associated depthmap.

Fig. 3. Example of an augmented spherical panorama obtained using a multi-baseline spherical camera.

problem (also known as visual odometry) is solved using a direct 3D image registration technique that provides accurate pose estimation [18]. Since this is a local optimisation approach it is assumed that the camera framerate is high (30Hz) and that interframe displacements are small ($\leq 2m$), meaning a maximum speed of $\sim 200km/h$. It is noted here that dense visual odometry is computationally efficient and locally very accurate [6] so it has been deemed unnecessary to perform costly bundle adjustment on local visibility windows (although this slightly improves the estimate, it makes timely scene acquisition practically infeasible).

In order to avoid reconstructing each sphere from the learning sequence and to obtained a graph with minimal redundancy, a robust statistical criteria is observed to choose where to reconstruct a sphere and add it to the graph. This allows compressing the original sequence of images to only few spherical images (see [18]).

Since a visual odometry approach is employed, drift is integrated over large scale trajectories (*e.g.* $\leq 1\%$). To overcome this problem, which might lead to inconsistencies in the graph, a spherical loop closure detection is employed [5]. This technique detects loop closures from the appearance of the images, using SIFT descriptors [15]. Loop closure detections allow to add new edges to the graph, corresponding to new constraints. The final graph can be globally optimised using a graph optimisation approach (*e.g.* [9]).

IV. Real-time localisation

It is considered that during online navigation, a current image $\boldsymbol{\mathcal{I}}$, captured by a generic camera (e.g. monocular, stereo or omnidirectional) and an initial guess $\widehat{\mathbf{T}}$ of the current camera position within the graph are available. This initial guess permits the extraction of the closest reference sphere $\boldsymbol{\mathcal{S}}^*$ from the graph. Contrary to non-spherical approaches, a sphere provides all viewing directions and therefore it is not necessary to consider the rotational distance (to ensure image overlap). The closest sphere is subsequently determined uniquely by translational distance. In particular this avoids choosing a reference sphere that has similar rotation but large translational difference which induces self occlusions of buildings and also differences in image resolution caused by distance (which affects direct registration methods).

Since a sphere provides all local information necessary for 6 *d.o.f.* localisation (geometric and photometric information), an accurate estimation of the pose is obtained by an efficient direct minimization:

$$\mathbf{e}(\mathbf{x}) = \boldsymbol{\mathcal{I}}\left(w(\widehat{\mathbf{T}}\mathbf{T}(\mathbf{x}); s(Z, \mathbf{q}_S^*))\right) - \boldsymbol{\mathcal{I}}_S^*\Big(s(Z_S, \mathbf{q}_S^*)\Big), \quad (5)$$

where $\mathbf{x}$ is the unknown 6 *d.o.f.* pose increment. The warping function $w(.)$ transfers the current image intensities onto the reference sphere pixels $\mathbf{q_S}$ through novel view synthesis [1], using depth information Z_S. The function $s(.)$ selects only informative pixels, *w.r.t.* the saliency map $\boldsymbol{\mathcal{W}}_S$ which is already pre-computed on the reference sphere [17]. This selection speeds up the tracking algorithm without neither degrading observability of 3D motion nor accuracy.

The error function $\mathbf{e}(\mathbf{x})$ is minimized using an iterative non-linear optimization (IRLS) detailed in Appendix VIII-A. The estimation is updated at each step by an homogeneous transformation:

$$\widehat{\mathbf{T}} \leftarrow \widehat{\mathbf{T}}\mathbf{T}(\mathbf{x}), \quad (6)$$

where $\widehat{\mathbf{T}}$ is the current pose estimate with respect to the closest reference sphere which is determined from the previous iterations up to time $t-1$.

A maximum amount of pre-computation is performed offline during the construction of the spheres (e.g. Jacobian matrices and saliency maps) allowing the online algorithm to be computationally very efficient: the camera pose can be estimated at high frame rate (*e.g.* 45 Hz for a current image of 800×600 pixels in size).

To farther improve performance, a coarse-to-fine optimization strategy is employed by using multi-resolution spheres (*e.g.* constructed by Gaussian filtering and sub-sampling [4]). The minimization begins at the lowest resolution and the result is used to initialize the next level repeatedly until the highest resolution is reached. This greatly improves the convergence domain/speed and some local minima can be avoided. Finally, to ensure robustness to illumination changes that occur between the reference images and the online camera's images, the method proposed in [19] is employed. This method combines both model-based tracking (*w.r.t.* the graph) and a non classic visual odometry approach, which greatly improve robustness to large illumination changes without necessitating the cost of estimating an illumination model.

V. Autonomous navigation

During autonomous navigation, the aim is to follow automatically a reference trajectory $\mathcal{U}$ generated locally around the learnt graph. The trajectory $\mathcal{U} = \{\mathbf{u}_1^*, \mathbf{u}_2^*, \ldots, \mathbf{u}_n^*\}$ contains n input vectors such that:

$$\mathbf{u}^* = \{x^*, y^*, \psi^*, U^*, \dot{\psi}^*\}, \quad (7)$$

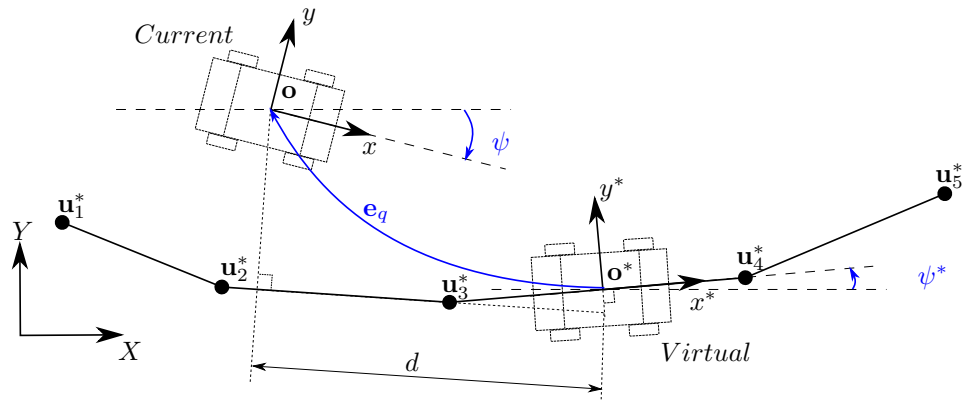


Fig. 4. Regulated error for visual servoing. The current vehicle position is projected onto the closest reference trajectory edge. A reference position is selected at a distance d, to generate longitudinal and angular errors $\{\mathbf{e}_q, e_\psi\}$.

where the point $\mathbf{o}^* = \{x^*, y^*\}$ is a desired position, ψ^* is the yaw angle, U^* is the longitudinal velocity and $\dot{\psi}^*$ is the desired angular velocity.

The control problem can be formulated as detailed in [3]. In the proposed case, a *virtual* vehicle is followed and used to generate an error in translation and orientation that can be regulated using state feedback. Longitudinal velocity is controlled using a proportional feedback on the longitudinal error and steering angle depends on yaw and transversal errors. These errors are obtained by projecting the current vehicle position onto the closest reference trajectory's edge (*cf.* Figure 4). The reference position is then selected by translating the projected position along the trajectory by a distance d.

The translation error between the reference point and the current position is then defined by:

$$\mathbf{e}_q = \begin{bmatrix} e_x \\ e_y \end{bmatrix} = \mathbf{R}_{\psi^*}^T (\mathbf{o} - \mathbf{o}^*) = \mathbf{R}_{\psi^*}^T \begin{bmatrix} x - x^* \\ y - y^* \end{bmatrix}, \quad (8)$$

where the rotation matrix of ψ^* can be written by:

$$\mathbf{R}_{\psi^*} \begin{bmatrix} \cos(\psi^*) & -\sin(\psi^*) \\ \sin(\psi^*) & \cos(\psi^*) \end{bmatrix}. \quad (9)$$

The angular error is directly defined by:

$$e_\psi = \psi - \psi^*, \quad (10)$$

and the control law, derived from [3] without the state feedback on the velocities is:

$$\begin{cases} U = U^* - k_x(|U^*| + \epsilon)e_x \\ \dot{\psi} = \dot{\psi}^* - k_y|U^*|e_y - k_\psi|U^*|\tan(e_\psi) \end{cases}, \quad (11)$$

where the gains k_x, k_y, k_ψ and ϵ are positive scalars.

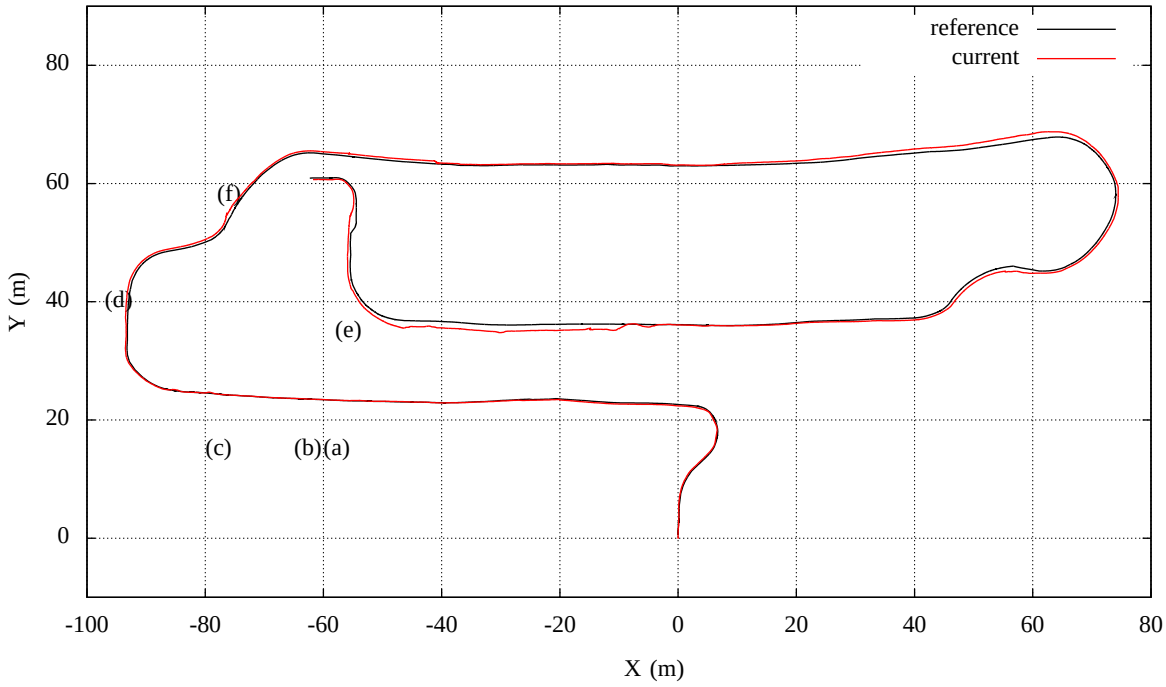
An accurate 6 *d.o.f.* vehicle localisation is computed using only a monocular camera as detailed in Section IV. The obtained 6 *d.o.f.* pose is then converted into a 3 *d.o.f.* position (x, y, ψ) in the vehicle plane, and the resulting error is regulated using equation (11).

Since the localisation algorithm provides a 45 Hz pose estimation, no additional sensors or filtering are used for the positioning and the vehicle is able to smoothly follow trajectories. To ensure safe experiments, a SICK laser is placed in the front of the vehicle and it is only employed to detect pedestrians and other vehicles. Longitudinal velocity is modulated with respect to the closest objects detected in the laser trace and the vehicle is stopped under a certain distance threshold (*cf.* Figure 5(e)).

A. Results

The mapping and re-localisation methods have been experimented in large scale environments. The following result shows autonomous navigation in the city centre of Clermont-Ferrand in France, obtained during the final experiments of the CityVIP project (*cf.* Section VII). A learning phase has been performed on a 490 meters trajectory, by manually driving a vehicle equipped with the spherical acquisition system. In order to ensure an admissible path for the online trajectory (*i.e.* without obstacles), the trajectory obtained during the learning phase was used as input for the online navigation: however the localisation method is capable of accurately localising a camera navigating within a different region in the graph, as it is demonstrated in [18]. The reference longitudinal velocity was set to 1.2m/s for the whole sequence.

Figure 5 shows the desired trajectory in black, and the trajectory followed autonomously by the vehicle in red. The vehicle starts at position (X=0,Y=0) and begins to move along Y axis. The experiment finishes at position (X=-62,Y=61). The vehicle was able to follow autonomously the whole sequence, using only the monocular camera for localisation. As it can be seen on figures (a),(b) and (c), the accuracy of the localisation method allows to navigate in narrowed corridors, whilst the employ of robust estimators ensure robustness to occlusions like pedestrians. Images (d) and (f) show the vehicle navigating in much larger areas (open place). This kind of environment has shown some limitations of vision based navigation. Since geometric information is far from the camera (building facade), accurate estimations of translations are degraded (infinite points are invariant to translations). These effects can be seen on the red trajectory around the landmark (e). However, this lack of precision could be overcomed using additional sensors, such as GPS and inertial measurements.



(a)

(b)

(c)

(d)

(e)

(f)

Fig. 5. Autonomous navigation. Top: A 490 meters trajectory followed autonomously, the reference trajectory is shown in black. The current trajectory is shown in red. (a),(b),(c),(d),(e),(f): images captured during the navigation.

VI. Conclusions

The approach described in this paper propose a fully automated mapping and re-localisation system, for autonomous navigation in urban environments. The mapping method allows to reconstruct dense visual maps of large scale 3D environments, using a spherical graph representation. It has been shown that this representation is capable of reproducing photometrically accurate views locally around a learnt graph. Reconstructed spheres acquired along a trajectory are used as input for a robust dense spherical tracking algorithm which estimates the spheres' positions.

During online navigation, an efficient direct registration technique is employed to accurately localise a monocular camera. The robustness of the localisation method has been validated in challenging urban sequences containing a lot of pedestrians and outliers. The proposed localisation system is only based on a standard monocular camera, no additional sensors are used.

Future work will aim at improving the database construction, by geo-referencing the spherical graph in a GIS (Geo-referenced Information System), which may contain higher level information such as free space. This geo-location will allow to use advanced path planning algorithms. To farther improve the autonomous navigation solution, it should be interesting to fuse vision based localisation, with GPS signal and inertial measurements.

VII. Acknowledgments

This work has been supported by ANR (French National Agency) CityVIP project under grant ANR-07_TSFA-013-01.

VIII. Appendix

A. Non-linear optimisation

The error function for the real-time tracking (5) is minimized using an iteratively re-weighted least squared non-linear minimization:

$$\mathcal{O}(\mathbf{x}) = \arg\min_{x} \rho(\mathbf{e}(\mathbf{x})), \tag{12}$$

by $\nabla\mathcal{O}(\mathbf{x})|_{\mathbf{x}=\tilde{\mathbf{x}}} = \mathbf{0}$, where ∇ is the gradient operator with respect to the unknown $\mathbf{x}$ defined in equation (2) assuming a global minimum is reached at $\mathbf{x} = \tilde{\mathbf{x}}$.

An efficient second order minimisation approach is employed [16], which allows to pre-compute most of the minimization parts directly on the reference image. In this case the unknown $\mathbf{x}$ is iteratively updated using a Gauss-Newton like optimization procedure:

$$\mathbf{x} = -(\mathbf{J}^T\mathbf{D}\mathbf{J})^{-1}\mathbf{J}^T\mathbf{D}\mathbf{e}(\mathbf{x}), \tag{13}$$

where ^{T} is the transposition operator, $\mathbf{J}^T\mathbf{D}\mathbf{J}$ is the robust Gauss-Newton Hessian approximation. $\mathbf{J}$ is the warping Jacobian matrix of dimension $n \times 6$. $\mathbf{D}$ is a diagonal weighting matrix of dimension $n \times n$ obtained by M-estimation [12] which rejects outliers such as occlusions and local illumination changes.

References

[1] S. Avidan and A. Shashua. Novel view synthesis in tensor space. *IEEE International Conference on Computer Vision and Pattern Recognition*, 0:1034, 1997.

[2] P. Baker, C. Fermuller, Y. Aloimonos, and R. Pless. A spherical eye from multiple cameras. *IEEE International Conference on Computer Vision and Pattern Recognition*, 1:576, 2001.

[3] S. Benhimane, E. Malis, P. Rives, and J.R. Azinheira. Vision-based control for car platooning using homography decomposition. In *IEEE International Conference on Robotics and Automation*, pages 2161 – 2166, april 2005.

[4] P. J. Burt and E. H. Adelson. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics*, 2:217–236, 1983.

[5] A. Chapoulie, P. Rives, and D. Filliat. A spherical representation for efficient visual loop closing. In *Proceedings of the 11th workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras*, 2011.

[6] A.I. Comport, E. Malis, and P. Rives. Real-time quadrifocal visual odometry. *The International Journal of Robotics Research*, 29(2-3):245–266, 2010.

[7] D. Craciun, N. Paparoditis, and F. Schmitt. Multi-view scans alignment for 3d spherical mosaicing in large-scale unstructured environments. *Computer Vision and Image Understanding*, 114(11):1248 – 1263, 2010. Special issue on Embedded Vision.

[8] C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems and practical applications. In *European Conference on Computer Vision*, pages 445–461, 2000.

[9] G. Grisetti, S. Grzonka, C. Stachniss, P. Pfaff, and W. Burgard. Efficient estimation of accurate maximum likelihood maps in 3d. In *IEEE International Conference on Intelligent Robots and Systems*, pages 3472 –3478, November 2007.

[10] K. Hammoudi, F. Dornaika, B. Soheilian, and N. Paparoditis. Generating raw polygons of street facades from a 2d urban map and terrestrial laser range data. In *SSSI Australasian Remote Sensing and Photogrammetry Conference*, 2010.

[11] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:328–341, 2008.

[12] P.J. Huber. *Robust Statistics*. New york, Wiley, 1981.

[13] J. Kopf, B. Chen, R. Szeliski, and M Cohen. Street slide: Browsing street level imagery. *ACM Transactions on Graphics*, 29(4):96:1 – 96:8, 2010.

[14] G. Krishnan and S.K. Nayar. Towards A True Spherical Camera. In *SPIE Human Vision and Electronic Imaging*, 2009.

[15] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[16] E. Malis. Improving vision-based control using efficient second-order minimization techniques. In *IEEE International Conference on Robotics and Automation*, volume 2, pages 1843–1848, 26-may 1, 2004.

[17] M. Meilland, A.I. Comport, and P. Rives. A spherical robot-centered representation for urban navigation. In *IEEE International Conference on Intelligent Robots and Systems*, pages 5196 –5201, 2010.

[18] M. Meilland, A.I. Comport, and P Rives. Dense visual mapping of large scale environments for real-time localisation. In *IEEE International Conference on Intelligent Robots and Systems*, pages 4242 –4248, sept. 2011.

[19] M. Meilland, A.I Comport, and P. Rives. Real-time dense visual tracking under large lighting variations. In *British Machine Vision Conference*, pages 45.1–45.11, 2011.

[20] S.K. Nayar. Catadioptric omnidirectional camera. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 482–, 1997.

[21] R. Szeliski. Image alignment and stitching: a tutorial. *Foundations and Trends in Computer Graphics and Vision*, 2(1):1–104, 2006.